R E P O R T R E S U M E S

ED 015 836 56 RE 001 034

A COMPARISON OF STANDARD AND VISUAL ADMINISTRATIONS OF THE CONCEPT MASTERY TEST.

BY- CURTIS, H.A. KROPP, R.P.

PUB DATE 61

GRANT OEG-7-08-075

EDRS PRICE MF-$0.25 HC-$0.68 15P.

DESCRIPTORS- *INSTRUCTIONAL INNOVATION, *TEST RELIABILITY, *TEST VALIDITY, PROJECTION EQUIPMENT, TESTING, STATISTICAL ANALYSIS, THE TERMAN CONCEPT MASTERY TEST FORM T,

A STUDY WAS CONDUCTED TO DETERMINE CHANGES IN SCORE CHARACTERISTICS RELATED TO VARIOUS CONDITIONS OF SPEEDEDNESS WHEN A TEST IS PRESENTED VISUALLY BY PROJECTING ONE ITEM AT A TIME. THIS METHOD OF STUDY IS POTENTIALLY USEFUL FOR PRESENTING TEST MATERIALS BY TELEVISION. THE TOTAL SCORE AND PART SCORE DATA REVEALED HIGH RELATIONSHIPS BETWEEN THE NORMAL AND THE PROJECTED PRESENTATION OF MATERIALS ADOPTED FROM THE TERMAN CONCEPT MASTERY TEST, FORM T, TO GRADUATE STUDENTS. THERE WAS TENTATIVE EVIDENCE THAT TESTS OF THIS TYPE, WHEN ADMINISTERED UNDER PACED, PROJECTED CONDITIONS, COULD BE SPEEDED GREATLY WITHOUT APPRECIABLY ALTERING TEST RELIABILITY AND VALIDITY. HOWEVER, SEPARATE NORMS WOULD NEED TO BE CONSTRUCTED FOR SCORES OBTAINED UNDER HIGHLY SPEEDED CONDITIONS WHICH PRODUCE SCORE DECREMENTS. IT WAS OBSERVED THAT PACED, VISUAL TESTING LED TO A LARGER NUMBER OF RESPONSES BY THE SUBJECT AND THAT THESE RESPONSES WERE GIVEN IN A SHORTER AMOUNT OF TIME AS COMPARED, IN BOTH CASES, TO PERFOMRANCE ON FORMS NORMALLY ADMINISTERED. THE INTRODUCTION OF AN INTEREXPOSURE INTERVAL DID NOT SEEM TO INCREASE SCORES. THE PACED, VISUAL TESTS SEEMED TO INDUCE A HIGHER LEVEL OF MOTIVATION ON THE PART OF SUBJECTS THAN DID THE TRADITIONAL PRESENTATION. THERE WAS INTENSE CONCENTRATION AND SUSTAINED PERFORMANCE BY EACH SUBJECT. (AUTHORS)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

✓ No BR no.
PA 510

ED015836

A COMPARISON OF STANDARD AND VISUAL ADMINISTRATIONS OF THE CONCEPT MASTERY TEST[1, 2]

H. A. Curtis and R. P. Kropp
Florida State University

The utilization of television as a medium of instruction is being widely and intensively studied, but its utilization as a testing medium has commanded very little attention. It would seem that televised instruction accompanied by televised testing would have particular value in situations in which the student body is widely dispersed. It would also seem that television may make distinct contributions to the field of psychometrics in that it offers sound and visual imagery (both still and motion) as well as operational control of the subject during the testing process. This is one of several reports dealing with basic problems involved in utilizing television as a testing medium.[3]

The limited size of the television screen prohibits presenting all items simultaneously as is done when a test is administered normally. A television presentation requires that items be presented serially, either individually or in small groups, for a pre-determined amount of time. This has the effect of "pacing" the subjects. Consequently, studies of the effects of item pacing are basic to television testing. In the study here reported, an analogue (slide projector and projection screen) was used to present the test materials to subjects thereby avoiding the expense and difficulty of a televised presentation.

RE001 034

---

[1] This investigation was supported by a grant from the United States Office of Education, National Defense Act of 1958, Title VII, Grant Number 7-08-075.

[2] To be published in AV Communication Review.

[3] Curtis, H. A. and R. P. Kropp, "A comparison of Scores Obtained by Administering a Test Normally and Visually." Journal of Experimental Education. In press.

Problem

The specific problem treated here is determining the effect on test scores caused by altering the exposure time per item when test items are presented one-at-a-time by projecting them onto a screen. Item exposure time was the independent variable except in two sets of data where the independent variable was a combination of intra-item and inter-item exposure time. Other variables believed to be relevant; e.g., arrangement of items according to difficulty, were held constant over all experimental and control administrations of the instrument.

Procedure

Experimental Test. The Terman Concept Mastery Test, Form T, was adapted for use in the study[4] Part I of the test consists of 115 word pairs; the subject is asked to indicate whether each pair is a pair of synonyms or antonyms. Part II consists of 75 analogies; the subject is asked to complete the analogy by choosing an alternative from the three presented for each analogy.

For experimental purposes, the test was divided into three forms after omitting item 115, Form A consisted of items 1, 4, 7, etc....; Form B, items 2, 5, 8, etc.; and Form C, items 3, 6, 9, etc. Separate booklets were prepared for each of these forms and they were labeled form TAC, TBC, and TCC, respectively. The items in each form were reproduced on slides, one item to a slide. These forms were labeled TAX, TBX, and TCX, respectively. Thus form TAC and TAX consisted of the same items, but the former was in normal booklet form and the latter consisted of slides for visual presentation.

---

[4] The authors wish to express their gratitude to the Psychological Corporation for permission to use this test experimentally.

The equivalence of these three forms was determined by administering under normal conditions all forms in random order to each of twenty-seven graduate students. Pairs of means, after correction for correlated performance, were not significantly different at the .05 level. Inter correlations between forms ranged from .784 to .848. Pairs of mean times spent on taking Part I, Part II, and the total test, respectively, were not different at the .05 level.

Subjects. The members of four classes of graduate students were used as subjects. All students for whom English was a second-language were eliminated before the data were analyzed. No subject appeared in more than one class. No data were collected about the comparability of the four groups prior to testing. However, the groups did not differ at the .05 level on the basis of their total rights score on the control form administered during the experimental testing.

Method. Each of the four groups was administered all three of the adapted forms of the CMT, two forms were projected onto a screen one item at a time, and the third form was administered under normal conditions. During the projected administrations of the forms, the instructions, sample items, etc., were also projected. Subjects were not informed of the length of time each item would be in view; however, they were informed that each item would be presented only one time. On the normal administration subjects were requested to indicate the times at which they completed Parts I and II.

An experimental condition is indicated in this manner "12-0, 24-0" which indicates that each item in Part I was exposed for 12 seconds, each item in Part II was exposed for 24 seconds, and there was no inter-exposure interval in either Part. Items in Part II received twice the exposure of items in Part I thereby preserving the approximate time distribution on both Parts revealed by the group on which the equivalency study was based.

Results

Total Scores. For each condition within each group, the mean score and standard deviation for rights, corrected rights, and omits were computed. Correlations and t tests, corrected for correlated performance, between rights, corrected rights, and omits were computed between conditions within each group. The means and standard deviations of the rights scores appear in Table 1. Correlations and t tests, corrected for correlated performance, appear in Table 2.

Table 1. Means and Standard Deviations for Total Scores

Group	Form	Condition	N	Rights	
				X	Sd
I	TAX	12-0, 24-0	16	36.56	9.3
	TBX	6-0, 12-0		35.00	8.5
	TCC	Normal		33.50	7.8
II	TBX	8-0, 16-0	19	33.84	11.35
	TCX	4-0, 8-0		27.74	13.12
	TAC	Normal		33.84	10.63
III	TBX	8-4, 16-8	12	37.33	7.15
	TAX	12-0, 24-0		36.83	7.67
	TCC	Normal		34.58	8.10
IV	TBX	6-4, 12-8	13	34.77	9.27
	TCX	10o0, 20-0		34.62	8.27
	TAC	Normal		32.46	9.77

Table 2. Correlations and t Tests between Mean Rights Scores

Group	N	Comparison	Rights	
			r	t
I	16	TAX - TBX	.875	1.345
		TAX - TCC	.878	2.666*
		TBX - TCC	.853	1.302
II	19	TCX - TBX	.790	-3.280*
		TCX - TAC	.889	-4.390*
		TBX - TAC	.861	.000
III	12	TAX - TBX	.794	- .347
		TAX - TCC	.714	1.249
		TBX - TCC	.890	2.469*
IV	13	TCX - TBX	.820	- .092
		TCX - TAC	.869	1.546
		TBX - TAC	.858	1.570

*Significant at .05 level.

The following observation might be made about the data appearing in these tables. It appears that intra-exposure intervals clustering around "8-0, 16-0", which is approximately equivalent to the time spent on the normal administration prorated over all items, yield about the same means as did the control administration. Usually the projected mean was slightly higher than the control mean except in the case of "4-0, 8-0", which constituted a drastic reduction in time. Generally the magnitude of intercorrelations between projected and control scores is as high or higher than that between pairs of projected administrations. In both conditions the introduction of an inter-exposure interval led to a very slight increase in mean score.

For three of the four groups, there were no differences between control and projected means after scores were corrected for guessing. The only significant differences between corrected means occurred in the group that was administered the projected test under the highly speeded "4-0, 8-0" condition.

The data reveal that the time spent on the control test by each of the four groups was less than that spent by a comparable group who supplied the data on which the equivalency of forms study was based. Inasmuch as the projected testing preceded the control testing, it might be that subjects unconsciously paced themselves to work at a faster than normal rate.

For each group, correlations were computed between the total time taken to complete the control test and total rights scores on the control and each experimental administration of test, separately. Except for one group, correlations were negative ranging from -.226 to -.465. On account of the small number of subjects in each group, the obtained correlations were not

significant; however, there seems to be a slight inverse relationship between normal speed of response and obtained score.

Although the relevant data are not presented here, analyses were run between groups of subjects formed on the basis of whether the projected test times were greater or lesser than their control test times. No differences significant at the .05 level were found between groups created in this way.

Part Scores. Analyses, of the same kind as those described above, on Part I and Part II, separately, reveal that the results described for the total test generally hold true for each part of the test despite the fact that Part I contains what are essentially true-false items and Part II contains multiple-choice items.

Discussion

This study was undertaken primarily to determine changes in score characteristics that are related to various conditions of speededness when a test is presented visually by projecting one item at a time. This method of test presentation merits study because it is potentially useful for presenting test materials by television.

The total score and part score data reveal high relationships between normal and projected presentation of standardized test materials. Of the various intra-exposure intervals only the "4-0, 8-0" condition, which is comparable to halving the normal test time, induced means that were statistically different from those obtained under normal conditions. But even under that condition, there was a high correlation between projected and normally obtained scores. Thus, there is tentative evidence that tests of this type when administered under paced, projected conditions can be speeded greatly without appreciably altering test reliability and validity.

However, separate norms would have to be constructed for scores obtained under highly speeded conditions which produce score decrements.

It was observed that the paced, visual testing led to a larger number of responses by the subject and that these responses were given a shorter amount of time as compared, in both cases, to performance on forms normally administered. Thus, paced, projected tests might permit presenting more items in a given amount of time than could be presented under normal conditions which would lead to a greater coverage of content and increased test validity.

The introduction of an inter-exposure interval did not seem to increase scores. It was assumed that the inter-exposure interval would reduce fatigue during the test. Even if the assumption is true, it is likely that two factors militated against a score increment. First, the items in Part I were synonym-antonym items and consequently it was simple for the subject to work during the inter-exposure interval and to indicate his response at the end of the interval. This notion is supported by the fact that the mean total scores on Part I are virtually the same for those conditions where the intra-exposure interval of one condition was equal to the inter- plus the intra-exposure interval of another condition. Second, each form was probably too short to cause an appreciable amount of fatigue in the subjects. Possibly the assumption would be supported by data collected from longer tests.

It would seem that if experimental data were available for projected conditions ranging downward from "8-0, 16-0", then it would be possible to determine the point of speededness at which the mean score decreases due to the diminished working time. It is necessary to establish such a point when a test is to be presented under paced, projected conditions because it indicates the maximum amount of speededness possible without appreciably altering the raw score distribution and consequently invalidating the norms.

Nevertheless, the correlation between scores collected under very speeded conditions and normally obtained scores are high and positive; therefore, suggesting the possibility of obtaining quite valid scores in substantially less time than is now taken under normal test administration conditions.

The reaction of subjects to the several testing conditions was closely observed to determine the apparent acceptability of projected tests and to obtain clues that might point to fruitful hypotheses concerning factors induced by the projection of tests. The investigators believe that the paced, visual tests induced a higher level of motivation on the part of subjects than did the traditional presentation. There was intense concentration and sustained performance by each subject. There was also the desire to discuss the testing experience with the investigators.

The most highly speeded presentation induced heated verbal hostility toward the investigators. Without further information about the subjects, it is difficult to explain definitely why the subjects behaved in this way. However, from their comments it seems that at least two factors prompted their reactions: (a) the feeling that the time was too short in which to give responses that fairly represented them, and (b) the feeling that the rules of fair play were violated in that they were asked to take the test and then were not given sufficient time to devote to it.

7/7/61 #385

Grant Number 7-08-075.00

Progress Report

1. Identification:

a. Department of Educational Research and Testing

School of Education

Florida State University

Tallahassee, Florida

b. Hazen A. Curtis and Russell P. Kropp, co-directors

c. Experimental Analyses of the Effects of Various Modes of Item Presentation on the Scores and Factorial Content of Tests Administered by Visual and Audio-Visual Means: A Program of Studies Basic to Television Testing.

d. January 1, 1961 to June 30, 1961.

e. Grant Number 7-08-075.00

2. Major Steps (Planning, Data Collection, Analysis, Recording and Reporting)

The following consists of reports of studies which were carried over from the previous reporting period and studies which were initiated during the period covered by this report.

1. A study of the effects on scores attributable to (a) intra-exposure interval, (b) inter-exposure interval, and (c) various combinations of (a) and (b) has been carried to the point of completion of preliminary analyses of data. The Terman Concept Mastery Test has been used as the experimental test and has been administered under normal conditions and two sets of experimental conditions to each of eight groups of college students. Exposure intervals (during which one item at a time is projected on a screen) have ranged systematically from an 80 per cent reduction of normal item-working time to three times normal working time. Preliminary data analyses reveal no significant F's for inter- and intra-exposure interval, and their interaction.

RE 001 034

A general account of the first part of this study has been accepted for publication by A-V Communications Review.[1] A copy is attached.

2. A study of the similarity of scores obtained by secondary school students from a vocational interest inventory--The Occupational Interest Inventory, Advanced--when administered under normal and various paced-projected conditions was undertaken and completed during the present reporting period. The experimental variable was intra-item exposure time. For one administration, the intra-exposure interval was given a duration equal to the average working time for the test when administered under normal conditions divided by the number of items. For the other experimental administration, the intra-exposure interval was two-thirds that described in the preceding sentence. Due to the ipsative nautre of the data coefficients of concordance were computed for each S between his ranking of vocational areas as revealed by the normal administration and his rankings obtained under each experimental condition, separately, and then all three simultaneously. Analyses revealed that (a) 75% to 80% of the obtained coefficients were equal to .80 or greater and (b) that coefficients between normally obtained rankings and experimental rankings obtained under the "normal" intra-exposure interval were, on the average, greater than, but not significantly so, than coefficients between normally obtained rankings and experimental rankings obtained when the intra-exposure interval was reduced to two-thirds its "normal" value. Generally the results indicate the feasibility of administering interest inventories such as this one under projected-paced conditions. Unfortunately, it was impossible to determine the predictive validities of the various sets of rankings. A copy of a full report of this study is attached.[2]

---

[1] Curtis, H.A., and Kropp, R.P. "A Comparison of Standard and Visual Administrations of the Concept Mastery Test." A-V Communications Review. In press.

[2] Stoker, H.W., Kropp, R.P., and Bashaw, W.L. "A Comparison of Scores Obtained Through Normal and Visual Administrations of the Occupational Interest Inventory." Available from Department of Educational Research and Testing, Florida State University, Tallahassee, Florida. Mimeographed.

3. A study was undertaken and completed that dealt with the comparisons between scores obtained from an achievement test when administered normally and under projected conditions. The test was an achievement test in political science and was constructed by the Department of Government of Florida State University. S's were registrants in two sections of an introductory course in political science which were taught by the same instructor. S's were divided into three groups and they were exposed to the following testing. One group was administered the test under normal conditions. A second group took the test under projected conditions -- one item was exposed at a time. The third group took the test under projected conditions but in addition when the item appeared on the screen it was read to them by means of tape recording which was synchronized with the projector. Under both projected conditions of testing a block of twenty-five items received a constant exposure interval and the remaining twenty-five items were given variable exposure items according to the "reading bulk" of each complete item. Analyses revealed: (a) no difference between control and experimental administrations; (b) no difference between projected silent and projected audio administrations nor between (c) constant and variable intra-item exposure interval. Again, results supported feasibility of the projected administration of achievements tests, and by implication supported televised administration of tests. A full account of this study has been accepted for publication and a copy is attached.[1]

4. A study conducted concomitantly with the one described immediately above was executed in the field of social studies to determine interrelationships between mode of testing (paper-and-pencil, projected silent, and projected audio) and certain intervening variables, particularly whether or not the sixth-grade students who were tested had long-term televised instruction in the social studies.[2]

---

[1] Kropp, R.P., King, F.J., Burr, W.L., and Board, J.G., "Differences in Achievement Test Scores Associated with Item Exposure Time and Audio Reinforcement." Journal of Educational Research. In press.

[2] Burr, W.L., Empirical Relationships Among Mode of Testing, Mode of Instruction and Reading Levels: in Sixth-grade social studies. Florida State University, 1961. Unpublished doctoral dissertation.

5. The study described here was undertaken to explore conditions under which personality and adjustment inventories might be administered under projected-silent conditions so as to obtain scores which are interchangeable with those obtained under typical paper-and-pencil conditions. A delinquency-proneness scale was fabricated by combining three somewhat empirically validated proneness scales which were published in the professional literature. Two groups of incarcerated delinquents and two groups of non-delinquent secondary school students, who were comparable on reading ability and age were administered the scale. One group of each responded to test items in scale which were projected one-at-a-time for a short interval on screen. Delinquents scores higher than nondelinquents on normal administration. However, groups tested under experimental conditions responded at about same level. Scores for non-delinquents, were higher on projected than on normal administration. Findings cast doubt on the interchangeability of testing methods for appraisals of this kind, and they simultaneously suggest new methods of testing for delinquency proneness. An account of this study has been accepted for publication and a copy is attached.[1]

6. An exploratory study which is underway at the present time deals with the possibility of validating items of the kind normally found in adjustment and personality inventories. The experimental test administration procedures, which are the subject to be studied under terms of the present grant, give rigid control over the presentation of the test thereby providing an opportunity to compare test behavior of subjects over a limited time and over known groups of items. As a consequence of this possibility, a study was devised to determine galvanic skin

---

[1] King, F.J., Kropp, R.P., Crawford, W.R., and Noyel, I.S., "Changes in Delinquency-Proneness Scores Associated with Paced-Projected and Normal Modes of Test Administration. Psychological Reports. In press.

reactions of subjects to adjustment inventory items that were presented under projected-paced-audio conditions as described earlier. All data have not been collected; however, preliminary analyses of that data already collected support the notion that the combination of the paced-projected-audio testing mode and the GSR protocols of subjects constitute a feasible method of determining whether a given item can potentially evoke an emotional reaction in subjects. Results of this study will certainly bear on more efficient techniques of item selection for such inventories.

7. Data will be collected an analyzed this summer bearing on the effect of various projected-paced testing conditions on scores from the control or validity scales of the Minnesota Multiphasic Personality Inventory. It is hypothesized that technique may lead to more valid scores. In short, the subject will give more valid responses, as revealed by changes on the validity scales, when he is denied time to reflect leisurely about his response to each of the items as he can do under the traditional paper-and-pencil testing conditions.

8. Data are now being analyzed that were collected to determine lower limits for intra-item exposure interval on paced-projected tests. Completed studies reveal that the paced-projected mode of testing permits reducing testing time drastically without altering the validity, reliability, or distribution of scores. The study deals with establishing modal response marking times for multiple-choice items, having different numbers of alternatives, when presented under projected conditions.

9. A study presently underway deals with establishing values for intra-item exposure intervals for tests which are neither purely power nor purely speeded tests; e.g., the subtests of the Primary Mental Abilities Test. Several subtests of the PMA were administered subjects under normal conditions except that time limits were ignored. The time each subject spent on each sub-test was recorded. Then, modal times were established for each subtest. Subjects were then re-tested under two sets of paced project conditions: (a) intra-exposure interval was

determined by distributing modal working time equally over all items, and (b) intra-exposure interval was determined as a function of item difficulty such that the summation of all exposure intervals equals modal working time. Data are being analyzed at the present time.

10. A study is underway dealing with the administration of a reading comprehension test-- Stanford Reading Comprehension Test, Advanced-- under three modes of administration--traditional, projected-silent, and projected-audio-- to determine if (a) traditional and projected-silent are interchangeable, (b) to determine if differences in scores achieved between projected-audio and either projected-silent or traditional constitute an index of reading retardation, and (c) if the factor structure of reading comprehension scores is altered by mode of testing. Data have been collected and the analysis is underway.

11. One of the early studies conducted under the grant revealed that arithmetic items that involved computation were difficult to present by the experimental testing procedure. A study is currently underway dealing with the similarity in validity between sets of scores on computational items when administered normally and sets of scores based on items designed to appraise the same arithmetic processes but which do not involve actual computational work by the subject during the test.

12. A study is being planned at the present time dealing with the establishment of threshold intra-exposure intervals for intelligence tests. Subjects will be tested individually and blocks of items will be presented at different exposure intervals. Orders will be randomly assigned subjects. Control data have been collected, but no experimental data have been collected.

13. Brief statement of next steps. Next steps include the completion of studies underway, undertaking and completing several new studies, preparing a final report, and preparing a proposal for the continuation of the general line of investigation which we believe will have great impact on the use of television in testing.

14. The only noteworthy departure from original plan has been the extension

of termination date of the grant from June 30, 1961 to September 30, 1961. Needless to say, the rejection of our proposal for a new grant under which to continue these studies has somewhat altered our long-term plans but has in no way caused us to deviate from what we originally proposed to do.

15. Summary of staff engaged in project.

Paid in whole from other funds:
H.A. Curtis--co-investigator
R.P. Kropp--co-investigator
H. Stoker--Assistant Prof. participates in studies
R.W. Willis--research assistant
W.L. Burr--research assistant

Paid in whole or in part from Federal funds:
F.J. King--research associate
W.L. Bashaw--research assistant
W.R. Crawford--research assistant
I.S. Moyal--research assistant
J.C. Beard--research assistant
M.E. Baker--research assistant
W. Pattu--consultant
Some part-time employees who perform clerical tasks and slide preparation.